Dependency Learning

Matthias C. M. Troffaes & Frank Coolen

Department of Mathematical Sciences
Durham University

9 September 2009

Outline

1. Fault Trees
   - Definition
   - Minimal Cut Set
   - Probability
2. Copulas
   - Definition
   - Dependency Model
   - Examples
3. Learning
   - Conjugate Analysis
   - Examples
   - Inverse Wishart

Outline

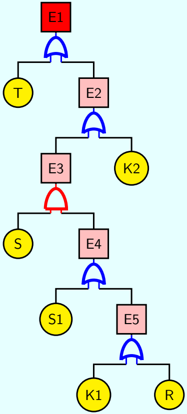
1. Fault Trees
   - Definition
   - Minimal Cut Set
   - Probability

2. Copulas
   - Definition
   - Dependency Model
   - Examples

3. Learning
   - Conjugate Analysis
   - Examples
   - Inverse Wishart

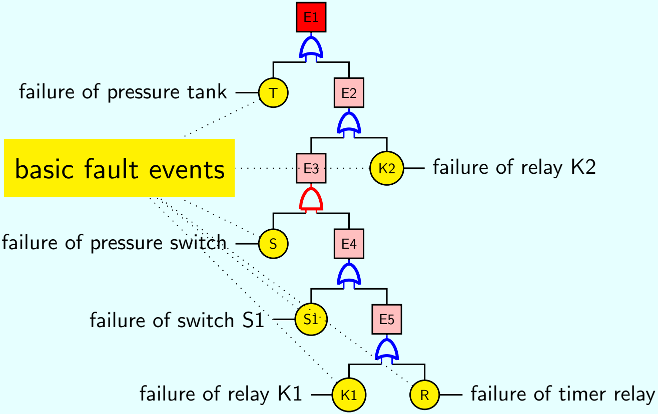
Fault Trees: Definition

structured representation of possible faults in a system



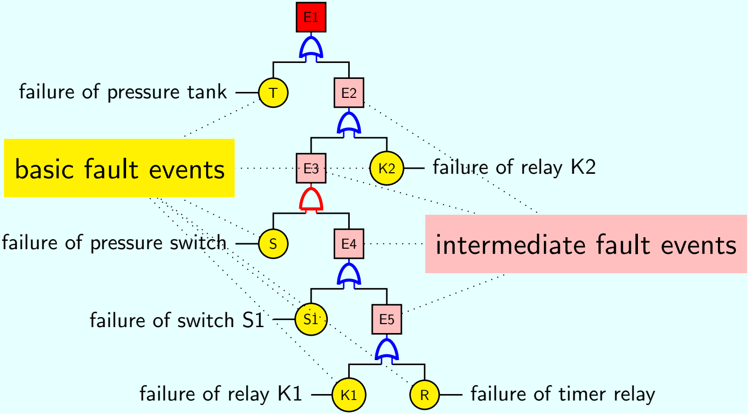
Fault Trees: Definition

structured representation of possible faults in a system



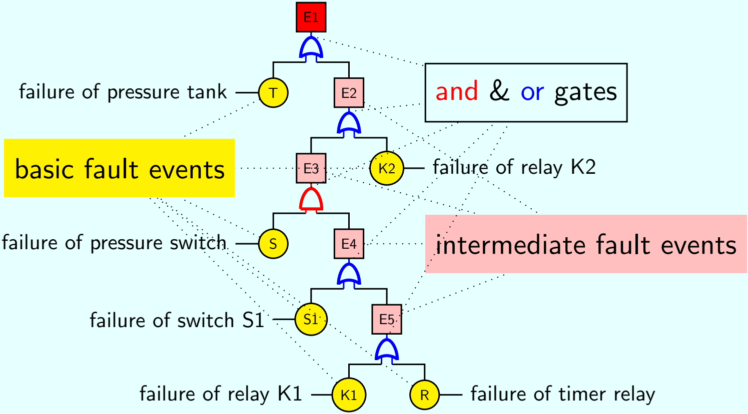
Fault Trees: Definition

structured representation of possible faults in a system



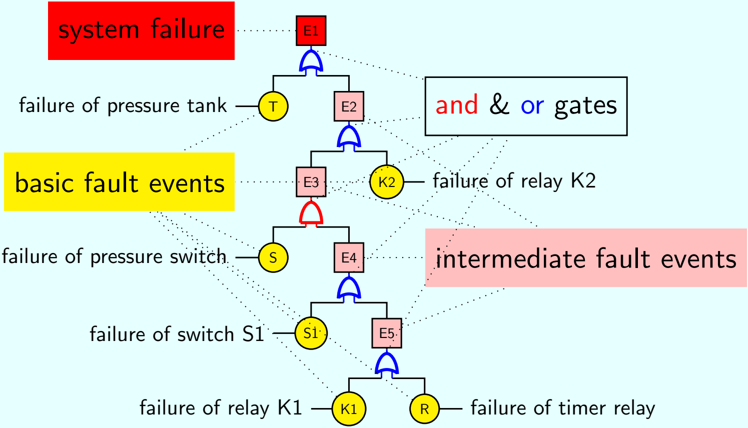
Fault Trees: Definition

structured representation of possible faults in a system



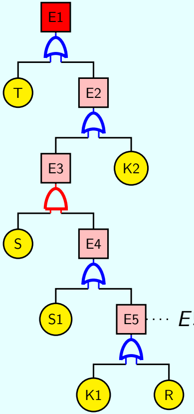
Fault Trees: Definition

structured representation of possible faults in a system



Fault Trees: Definition

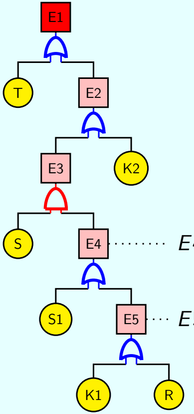
structured representation of possible faults in a system



$E5 = K1 + R$

Fault Trees: Definition

structured representation of possible faults in a system



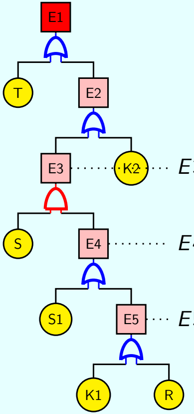
$$E4 = S1 + E5 = S1 + K1 + R$$

$$E5 = K1 + R$$

Durham University

Fault Trees: Definition

structured representation of possible faults in a system



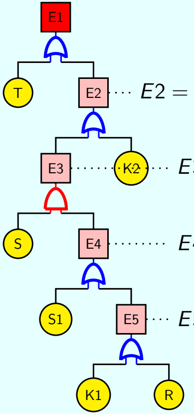
$E3 = S \cdot E4 = S \cdot (S1 + K1 + R)$

$E4 = S1 + E5 = S1 + K1 + R$

$E5 = K1 + R$

Fault Trees: Definition

structured representation of possible faults in a system



$$E2 = E3 + K2 = K2 + S \cdot (S1 + K1 + R)$$

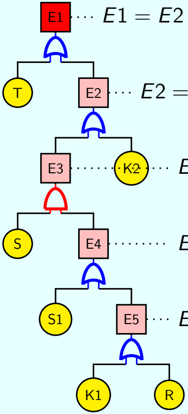
$$E3 = S \cdot E4 = S \cdot (S1 + K1 + R)$$

$$E4 = S1 + E5 = S1 + K1 + R$$

$$E5 = K1 + R$$

Fault Trees: Definition

structured representation of possible faults in a system



$$E1 = E2 + T = T + K2 + S \cdot (S1 + K1 + R)$$

$$E2 = E3 + K2 = K2 + S \cdot (S1 + K1 + R)$$

$$E3 = S \cdot E4 = S \cdot (S1 + K1 + R)$$

$$E4 = S1 + E5 = S1 + K1 + R$$

$$E5 = K1 + R$$

Fault Trees: Minimal Cut and Path Set

$$E1 = T + K2 + S \cdot (S1 + K1 + R)$$

- fault tree represents boolean expression
- two standard ways of rewriting these expressions

Fault Trees: Minimal Cut and Path Set

$$E1 = T + K2 + S \cdot (S1 + K1 + R)$$

- fault tree represents boolean expression
- two standard ways of rewriting these expressions
  - **minimal cut set**: which combination of component failures *causes* system failure?

$$E1 =$$

Fault Trees: Minimal Cut and Path Set

$$E1 = T + K2 + S \cdot (S1 + K1 + R)$$

- fault tree represents boolean expression
- two standard ways of rewriting these expressions
  - **minimal cut set**: which combination of component failures *causes* system failure?

$$E1 = \boxed{T}$$

Fault Trees: Minimal Cut and Path Set

$$E1 = T + K2 + S \cdot (S1 + K1 + R)$$

- fault tree represents boolean expression
- two standard ways of rewriting these expressions
  - **minimal cut set**: which combination of component failures *causes* system failure?

$$E1 = \boxed{T} + \boxed{K2}$$

Fault Trees: Minimal Cut and Path Set

$$E1 = T + K2 + S \cdot (S1 + K1 + R)$$

- fault tree represents boolean expression
- two standard ways of rewriting these expressions
  - **minimal cut set**: which combination of component failures *causes* system failure?

$$E1 = \boxed{T} + \boxed{K2} + \boxed{S \cdot S1}$$

Fault Trees: Minimal Cut and Path Set

$$E1 = T + K2 + S \cdot (S1 + K1 + R)$$

- fault tree represents boolean expression
- two standard ways of rewriting these expressions
  - **minimal cut set**: which combination of component failures *causes* system failure?

$$E1 = \boxed{T} + \boxed{K2} + \boxed{S \cdot S1} + \boxed{S \cdot K1}$$

Fault Trees: Minimal Cut and Path Set

$$E1 = T + K2 + S \cdot (S1 + K1 + R)$$

- fault tree represents boolean expression
- two standard ways of rewriting these expressions
    - **minimal cut set**: which combination of component failures *causes* system failure?

$$E1 = \boxed{T} + \boxed{K2} + \boxed{S \cdot S1} + \boxed{S \cdot K1} + \boxed{S \cdot R}$$

Fault Trees: Minimal Cut and Path Set

$$E1 = T + K2 + S \cdot (S1 + K1 + R)$$

- fault tree represents boolean expression
- two standard ways of rewriting these expressions
  - **minimal cut set**: which combination of component failures *causes* system failure?

$$E1 = \boxed{T} + \boxed{K2} + \boxed{S \cdot S1} + \boxed{S \cdot K1} + \boxed{S \cdot R}$$

  - **minimal path set**: which combination of component non-failures *prevents* system failure?

$$E1' =$$

Fault Trees: Minimal Cut and Path Set

$$E1 = T + K2 + S \cdot (S1 + K1 + R)$$

- fault tree represents boolean expression
- two standard ways of rewriting these expressions
  - **minimal cut set**: which combination of component failures *causes* system failure?

$$E1 = T + K2 + S \cdot S1 + S \cdot K1 + S \cdot R$$

  - **minimal path set**: which combination of component non-failures *prevents* system failure?

$$E1' = T' \cdot K2' \cdot S'$$

Fault Trees: Minimal Cut and Path Set

$$E1 = T + K2 + S \cdot (S1 + K1 + R)$$

- fault tree represents boolean expression
- two standard ways of rewriting these expressions
  - **minimal cut set**: which combination of component failures *causes* system failure?

$$E1 = T + K2 + S \cdot S1 + S \cdot K1 + S \cdot R$$

  - **minimal path set**: which combination of component non-failures *prevents* system failure?

$$E1' = T' \cdot K2' \cdot S' + T' \cdot K1' \cdot K2' \cdot S1' \cdot R'$$

Fault Trees: Probability of System Failure

if

- component failure probabilities are known and small
- components are independent

Fault Trees: Probability of System Failure

if

- component failure probabilities are known and small
- components are independent

then, use minimal cut sets, and plug in component failure probabilities (Vesley *et al.*, 1981, VIII-14)

$$P(E1) \approx P(T) + P(K2) + P(S)P(S1) + P(S)P(K1) + P(S)P(R)$$

rare event approximation

Fault Trees: Probability of System Failure

if

- component upper failure probabilities are known
- dependence between components is unknown

Fault Trees: Probability of System Failure

if

- component upper failure probabilities are known
- dependence between components is unknown

then, we can still write (Hoeffding, 1940; Walley, 1991, §2.7.4(d))

$$\overline{P}(E1) \leq \overline{P}(T) + \overline{P}(K2) + \overline{P}(S \cdot S1) + \overline{P}(S \cdot K1) + \overline{P}(S \cdot R)$$

and

$$\overline{P}(A \cdot B) \leq \min\{\overline{P}(A), \overline{P}(B)\}$$

Fault Trees: Probability of System Failure

if

- component upper failure probabilities are known
- dependence between components is unknown

then, we can still write (Hoeffding, 1940; Walley, 1991, §2.7.4(d))

$$\overline{P}(E1) \leq \overline{P}(T) + \overline{P}(K2) + \overline{P}(S \cdot S1) + \overline{P}(S \cdot K1) + \overline{P}(S \cdot R)$$

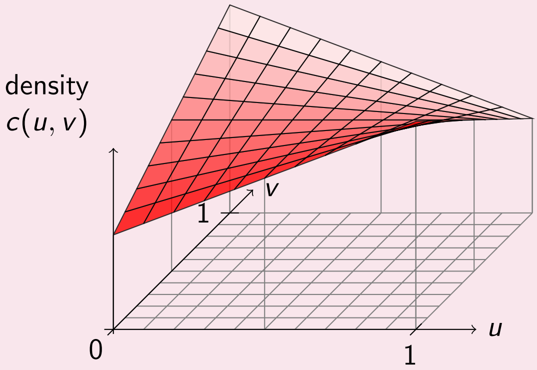
and

$$\overline{P}(A \cdot B) \leq \min\{\overline{P}(A), \overline{P}(B)\}$$

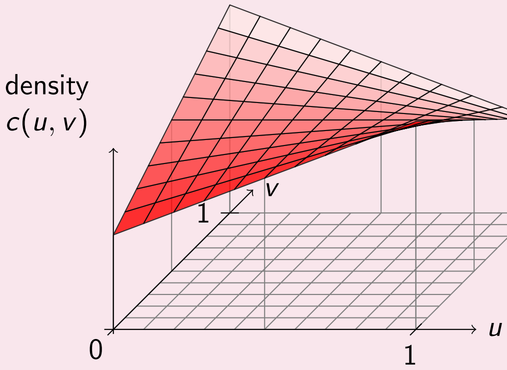
If we have joint data about two components, can we do better for the bound on $\overline{P}(A \cdot B)$?

Outline

Copulas: Definition



Copulas: Definition



- bivariate cumulative distribution $C(u, v)$ on unit square
- uniform marginals

Copulas: Dependency Model

Theorem (Sklar's Theorem (1959))

For any continuous bivariate cumulative distribution $H(x, y)$ on the real plane with marginal cumulative distributions $F(x)$ and $G(y)$, there is a copula $C(u, v)$ such that

$$H(x, y) = C(F(x), G(y))$$

Copulas: Example — Product Copula

Combines

- any marginals

into

- independent joint.

Copulas: Example — Product Copula

Combines

- any marginals

into

- independent joint.

$$C(u, v) = uv$$

$$\text{density } c(u, v) = 1$$

Copulas: Example — Gaussian Copula

Combines

- Gaussian marginals

into

- bivariate Gaussian joint with given correlation.

Copulas: Example — Gaussian Copula

Combines

- Gaussian marginals

into

- bivariate Gaussian joint with given correlation.

$$\text{density } c_\rho(\Phi(x), \Phi(y)) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right)$$

$(-1 < \rho < 1)$

Copulas: Example — Farlie-Gumbel-Morgenstern (FGM)

Polynomial perturbation of the product copula.

Copulas: Example — Farlie-Gumbel-Morgenstern (FGM)

Polynomial perturbation of the product copula.

$$C_\theta(u, v) = uv(1 + \theta(1 - u)(1 - v))$$
$$\text{density } c_\theta(u, v) = 1 + \theta(1 - 2u)(1 - 2v)$$

$(-1 \leq \theta \leq 1)$

Copulas: Example — Others

Many more copulas are studied in the literature!

Copulas: Relevance for Fault Trees

If we have joint data about two components, can we do better for the bound on $\overline{P}(A \cdot B)$?

Copulas: Relevance for Fault Trees

If we have joint data about two components, can we do better for the bound on $\overline{P}(A \cdot B)$?

Typical situation:

- $A = X \leq x$ and $B = Y \leq y$
- marginals $F(x)$ and $G(y)$ well known

Copulas: Relevance for Fault Trees

If we have joint data about two components, can we do better for the bound on $\overline{P}(A \cdot B)$?

Typical situation:

- $A = X \leq x$ and $B = Y \leq y$
- marginals $F(x)$ and $G(y)$ well known

so to know

$$P(A \cdot B) = H(x, y) = C(F(x), G(y)) = C(P(A), P(B))$$

it suffices to know the copula $C(u, v)$ for the joint $H(x, y)$!

Outline

Learning Copulas: Bayesian Approach

If we have joint data about two components, can we learn about $C(u, v)$?

Learning Copulas: Bayesian Approach

If we have joint data about two components, can we learn about $C(u,v)$?

If we have a parametric family of copulas

$$\{c_\alpha(u,v)\colon \alpha\}$$

Learning Copulas: Bayesian Approach

If we have joint data about two components, can we learn about $C(u, v)$?

If we have a parametric family of copulas

$$\{c_\alpha(u, v) \colon \alpha\}$$

with corresponding likelihood for data $\vec{d} = (x_1, y_1), \dots, (x_n, y_n)$

$$h(\vec{d}|\alpha) = \prod_{i=1}^{n} c_\alpha(F(x_i), G(y_i)) f(x_i) g(y_i) \propto \prod_{i=1}^{n} c_\alpha(F(x_i), G(y_i))$$

Learning Copulas: Bayesian Approach

If we have joint data about two components, can we learn about $C(u, v)$?

If we have a parametric family of copulas

$$\{c_\alpha(u, v) \colon \alpha\}$$

with corresponding likelihood for data $\vec{d} = (x_1, y_1), \dots, (x_n, y_n)$

$$h(\vec{d}|\alpha) = \prod_{i=1}^{n} c_\alpha(F(x_i), G(y_i)) f(x_i) g(y_i) \propto \prod_{i=1}^{n} c_\alpha(F(x_i), G(y_i))$$

can we then find a family of conjugate priors on α?

Learning Copulas: Conjugate for FGM Copula

$$h(\vec{d}|\rho) \propto \prod_{i=1}^{n}(1+\theta(1-2F(x_i))(1-2G(y_i))) = \prod_{i=1}^{n}(1+\theta u_i v_i)$$

(with $u_i = 1 - 2F(x_i)$ and $v_i = 1 - 2G(y_i)$)

Learning Copulas: Conjugate for FGM Copula

$$h(\vec{d}|\rho) \propto \prod_{i=1}^{n}(1+\theta(1-2F(x_i))(1-2G(y_i))) = \prod_{i=1}^{n}(1+\theta u_i v_i)$$

(with $u_i = 1 - 2F(x_i)$ and $v_i = 1 - 2G(y_i)$) has conjugate priors

$$p(\theta|\nu, a_1, \dots, a_\nu) \propto \prod_{k=1}^{\nu}(1 + a_k\theta)$$

Learning Copulas: Conjugate for FGM Copula

$$h(\vec{d}|\rho) \propto \prod_{i=1}^{n}(1+\theta(1-2F(x_i))(1-2G(y_i))) = \prod_{i=1}^{n}(1+\theta u_i v_i)$$

(with $u_i = 1-2F(x_i)$ and $v_i = 1-2G(y_i)$) has conjugate priors

$$p(\theta|\nu, a_1, \dots, a_\nu) \propto \prod_{k=1}^{\nu}(1+a_k\theta)$$

with updating rule

$$\nu \to \nu + n \qquad a_k \to a_k \text{ for } k \leq \nu$$
$$a_{\nu+i} = u_i v_i \text{ for } 1 \leq i \leq n$$

Durham University

stuck!

stuck!

Challenges:

- models for sets of polynomial distributions on $[-1, 1]$?
- reduce an infinite dimensional parameter set?
    - lower bound on variance?

Learning Copulas: Conjugate for Gaussian Copula

$$h(\vec{d}|\rho) \propto \frac{1}{\sqrt{1-\rho^2}^n} \exp\left(-\frac{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - 2\rho\sum_{i=1}^n x_i y_i}{2(1-\rho^2)}\right)$$

Learning Copulas: Conjugate for Gaussian Copula

$$h(\vec{d}|\rho) \propto \frac{1}{\sqrt{1-\rho^2}^n} \exp\left(-\frac{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - 2\rho \sum_{i=1}^n x_i y_i}{2(1-\rho^2)}\right)$$

has possible class of conjugate priors $p(\rho|\nu, \alpha, \beta)$:

$$p(\rho|\nu, \alpha, \beta) \propto (1-\rho^2)^{-\nu/2} \exp\left(-\frac{\alpha - 2\beta\rho}{1-\rho^2}\right)$$

Learning Copulas: Conjugate for Gaussian Copula

$$h(\vec{d}|\rho) \propto \frac{1}{\sqrt{1-\rho^2}^n} \exp\left(-\frac{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - 2\rho \sum_{i=1}^n x_i y_i}{2(1-\rho^2)}\right)$$

has possible class of conjugate priors $p(\rho|\nu, \alpha, \beta)$:

$$p(\rho|\nu, \alpha, \beta) \propto (1-\rho^2)^{-\nu/2} \exp\left(-\frac{\alpha - 2\beta\rho}{1-\rho^2}\right)$$

with updating rule

$$\nu \to \nu + n \qquad \alpha \to \alpha + \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2$$

$$\beta \to \beta + \sum_{i=1}^n x_i y_i$$

Learning Copulas: Conjugate for Gaussian Copula

$$h(\vec{d}|\rho) \propto \frac{1}{\sqrt{1-\rho^2}^n} \exp\left(-\frac{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - 2\rho\sum_{i=1}^n x_i y_i}{2(1-\rho^2)}\right)$$

has possible class of conjugate priors $p(\rho|\nu,\alpha,\beta)$:

$$p(\rho|\nu,\alpha,\beta) \propto (1-\rho^2)^{-\nu/2} \exp\left(-\frac{\alpha - 2\beta\rho}{1-\rho^2}\right) = \text{Troffaes distribution?}$$

with updating rule

$$\nu \to \nu + n \qquad \alpha \to \alpha + \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2$$

$$\beta \to \beta + \sum_{i=1}^n x_i y_i$$

stuck!

stuck!

Challenges:

- study this unknown distribution

Side Track: Conjugate for Gaussian Bivariate

(known mean, unknown covariance)

$$h(\vec{d}|\Sigma) \propto \prod_{i=1}^{n} N\left(\mu = \begin{bmatrix}0\\0\end{bmatrix}, \Sigma = \begin{bmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{bmatrix}\right)$$

has as conjugate prior the inverse-Wishart distribution

$$\Sigma \sim W^{-1}(\nu, \Psi)$$

with updating rule

$$\nu \to \nu + n \qquad \Psi \to \Psi + \begin{bmatrix}\sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i\\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} y_i^2\end{bmatrix}$$

Side Track: Conjugate for Gaussian Bivariate

Expectation for Σ, after reparametrisation

$$E(\Sigma|\nu + 3, \nu S) = S$$

Side Track: Conjugate for Gaussian Bivariate

Expectation for Σ, after reparametrisation

$$E(\Sigma|\nu+3, \nu S) = S$$

prior near-ignorance about correlation?

$$\left\{ W^{-1}(\nu+3, \nu S_{\sigma_X,\sigma_Y,\rho}) : -1 < \rho < 1 \right\}$$

with

$$S_{\sigma_X,\sigma_Y,\rho} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

Side Track: Conjugate for Gaussian Bivariate

Posterior expectation turns out to be

$$E(\Sigma|\vec{d}, \nu + 3, \nu S_{\sigma_X,\sigma_Y,\rho}) = S_{\sigma'_X,\sigma'_Y,\rho'}$$

with

$$\sigma'_X = \sqrt{\frac{\nu\sigma_X^2 + \sum_{i=1}^n x_i^2}{\nu + n}} \qquad \sigma'_Y = \sqrt{\frac{\nu\sigma_Y^2 + \sum_{i=1}^n y_i^2}{\nu + n}}$$

$$\rho' = \frac{\sigma_X\sigma_Y\left(\sum_{i=1}^n \frac{x_i y_i}{\sigma_X\sigma_Y} + \rho\nu\right)}{\sigma'_X\sigma'_Y(n+\nu)}$$

Side Track: Conjugate for Gaussian Bivariate

Imprecision?

$$[\underline{\rho'}, \overline{\rho'}] = \left[\frac{\sigma_X \sigma_Y \left(\sum_{i=1}^n \frac{x_i y_i}{\sigma_X \sigma_Y} - \nu \right)}{\sigma'_X \sigma'_Y (n+\nu)}, \frac{\sigma_X \sigma_Y \left(\sum_{i=1}^n \frac{x_i y_i}{\sigma_X \sigma_Y} + \nu \right)}{\sigma'_X \sigma'_Y (n+\nu)} \right]$$

or, if $\sigma_X = \sigma_Y = 1$ and sample variance agrees with prior variance

$$[\underline{\rho'}, \overline{\rho'}] = \left[\frac{\sum_{i=1}^n x_i y_i - \nu}{n+\nu}, \frac{\sum_{i=1}^n x_i y_i + \nu}{n+\nu} \right]$$

not stuck! :-)

not stuck! :-)

Challenges:

- non-Gaussian marginals?
- other families of copulas?

Conclusion

- bivariate Gaussian: joint data can be incorporated into the model quite easily, even accounting for prior ignorance
- Bayesian learning about dependencies via copulas is non-trivial: major challenges!
  - conjugate priors are easily found, but...
  - ways to reduce dimensionality? (imprecise probability has an advantage here!)
  - new distributions arise, begging to be studied
- also updating the (precise) marginals in the model can make the mathematics easier

Conclusion

- bivariate Gaussian: joint data can be incorporated into the model quite easily, even accounting for prior ignorance
- Bayesian learning about dependencies via copulas is non-trivial: major challenges!
  - conjugate priors are easily found, but...
  - ways to reduce dimensionality? (imprecise probability has an advantage here!)
  - new distributions arise, begging to be studied
- also updating the (precise) marginals in the model can make the mathematics easier

Thanks for your attention!

questions? comments? discussion?